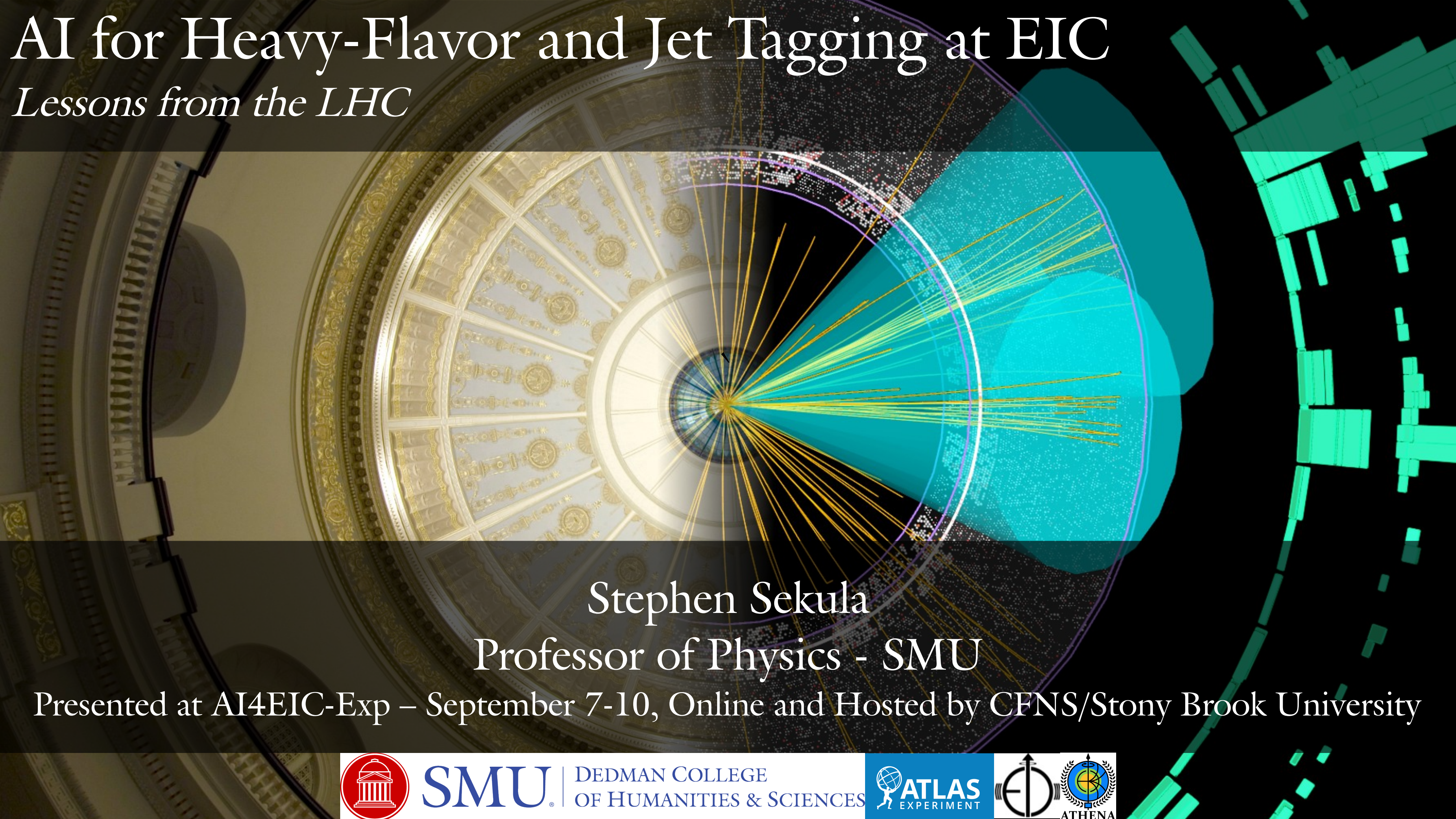


AI for Heavy-Flavor and Jet Tagging at EIC

Lessons from the LHC

Stephen Sekula

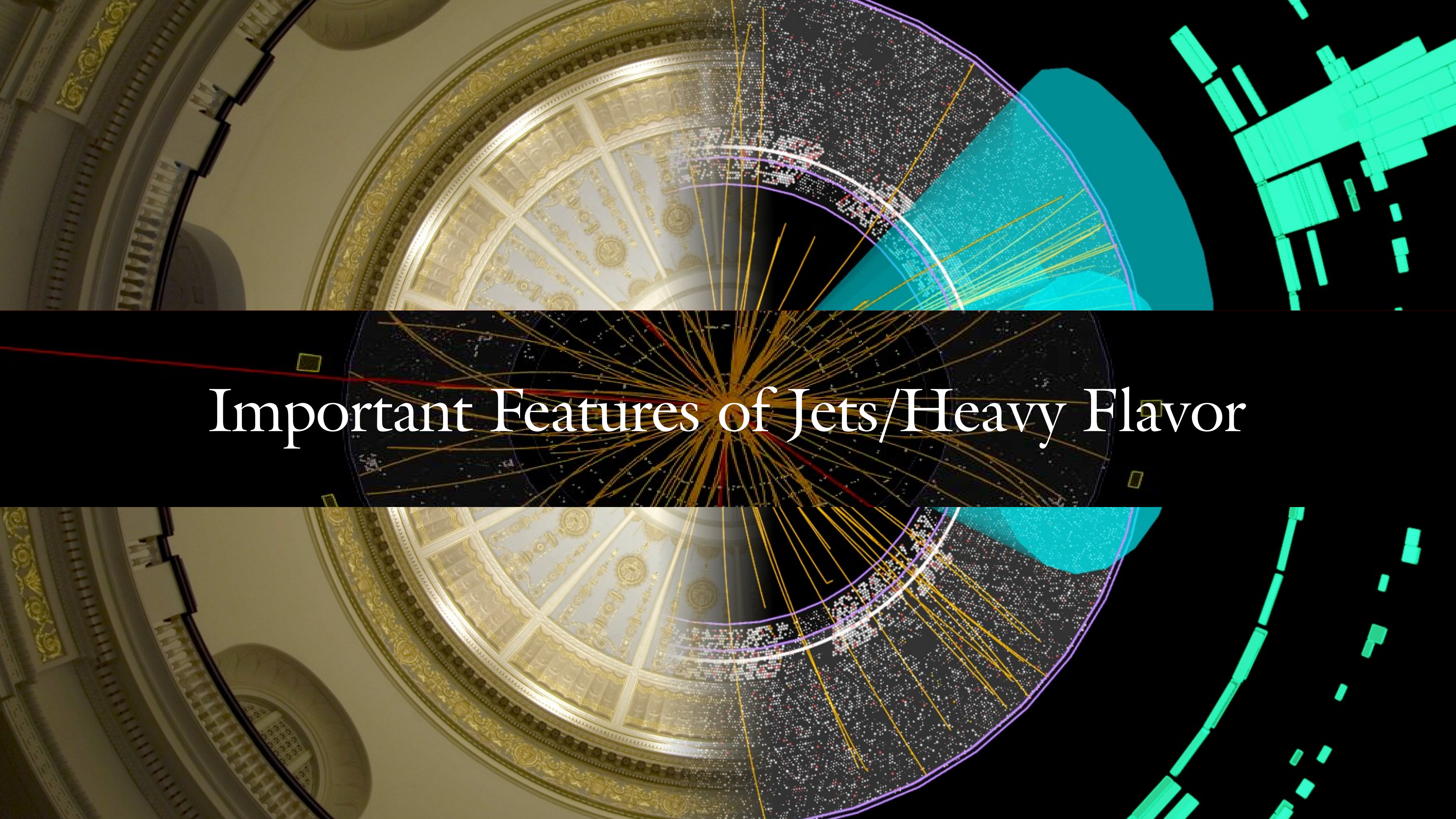
Professor of Physics - SMU

Presented at AI4EIC-Exp – September 7-10, Online and Hosted by CFNS/Stony Brook University

SMU | DEDMAN COLLEGE OF HUMANITIES & SCIENCES

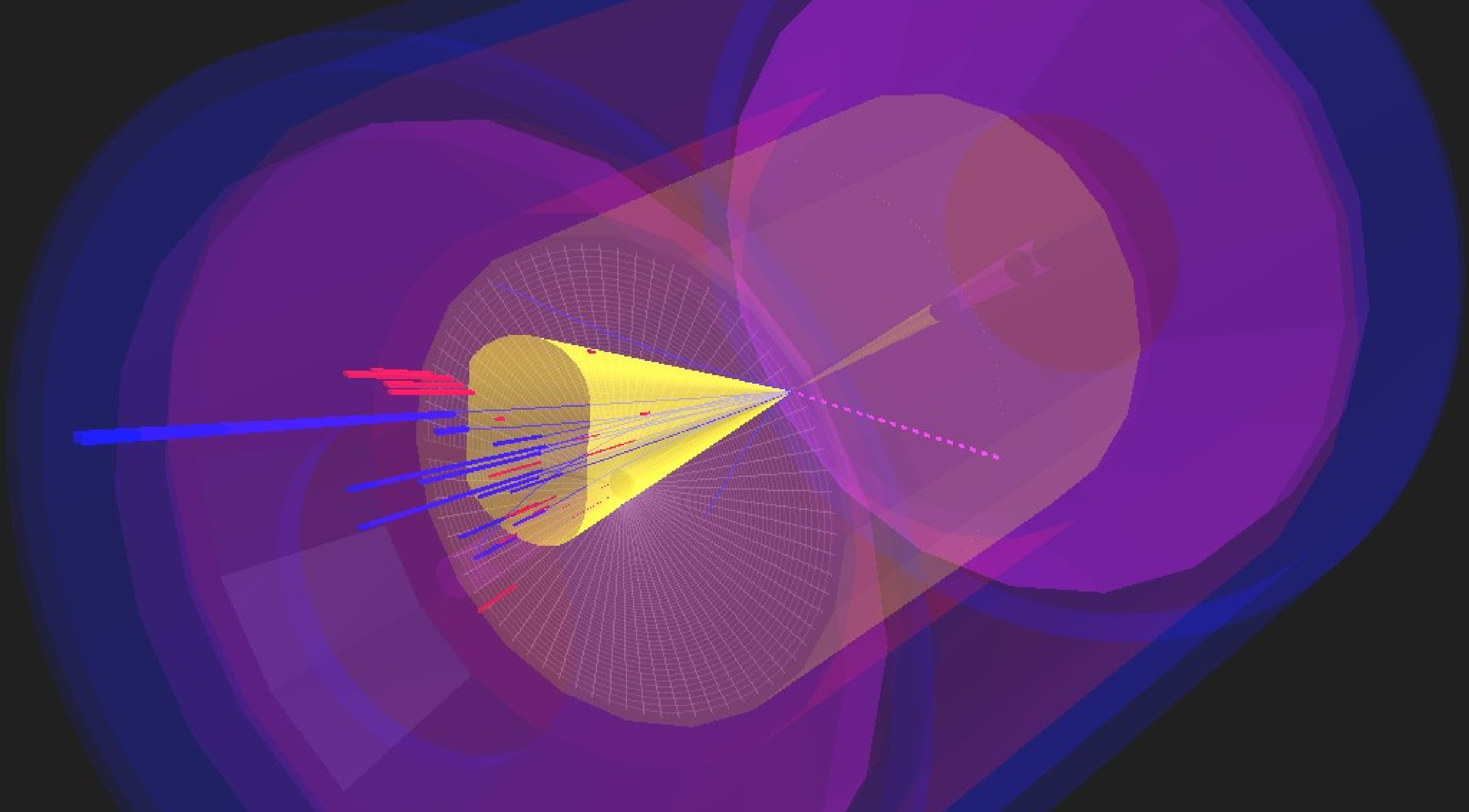
ATLAS EXPERIMENT

ATHENA



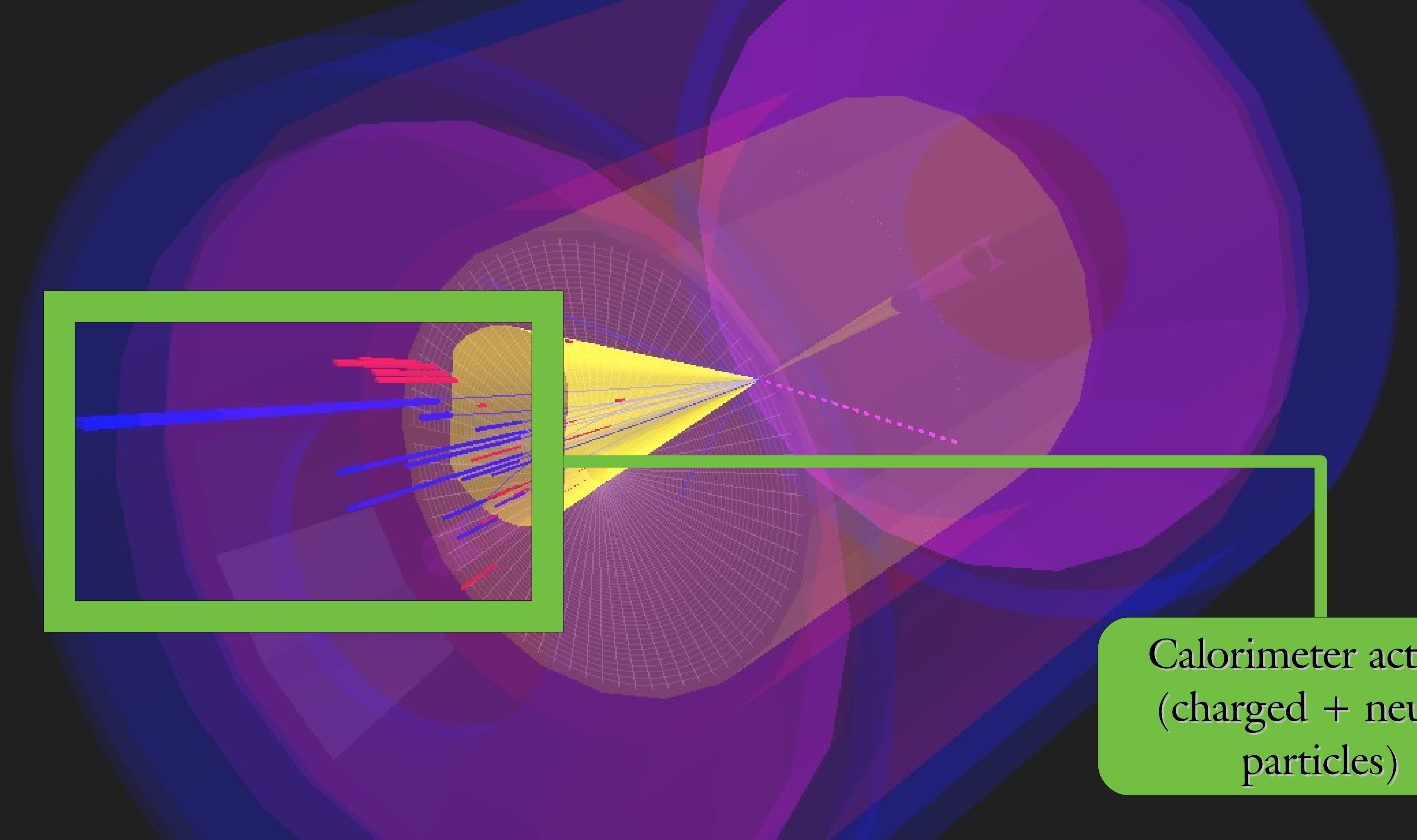
Important Features of Jets/Heavy Flavor

Jet and Heavy Flavor production will be common at EIC…



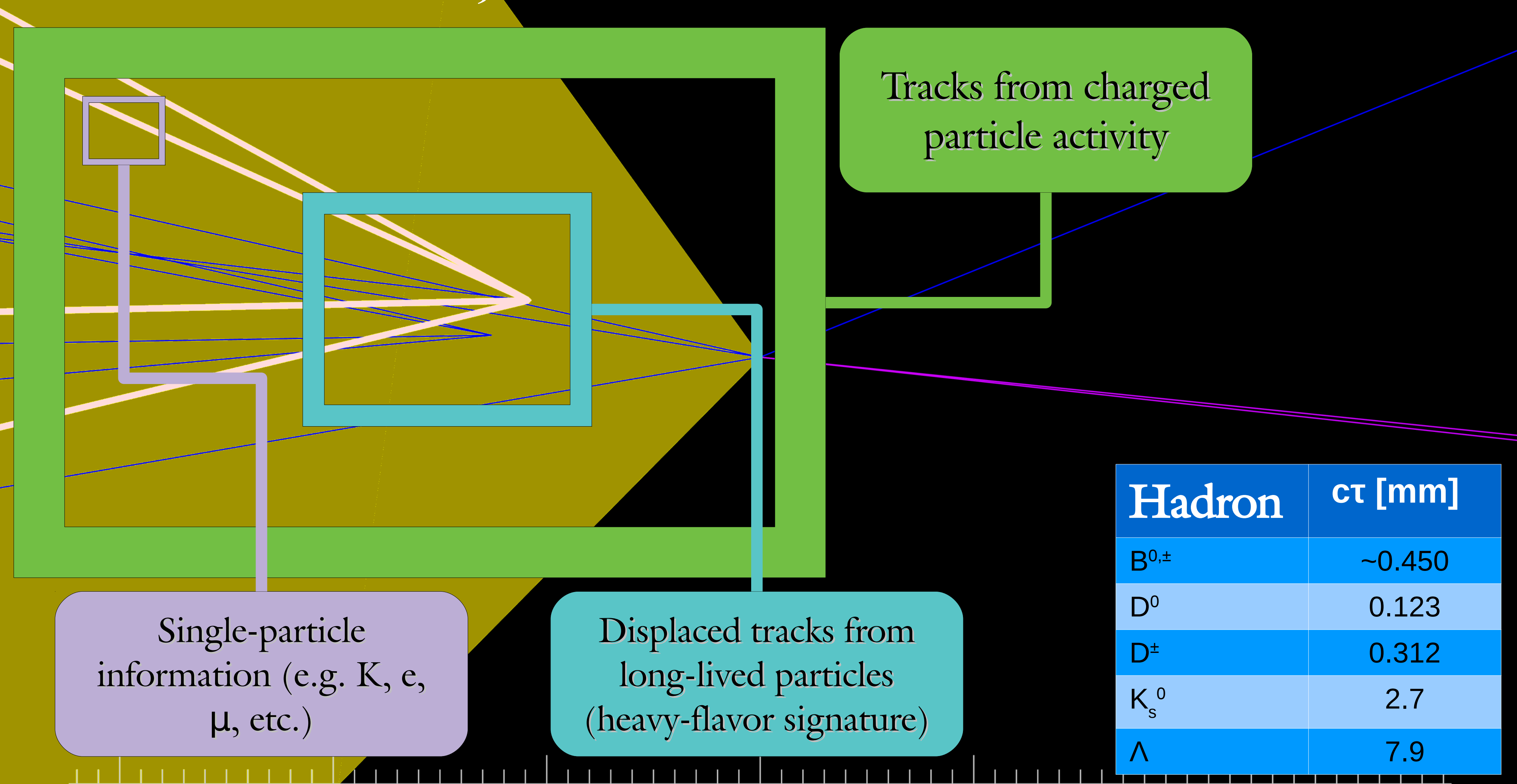
[Fast simulation of charged-current deep-inelastic scatter at EIC, 10x275 ep configuration, Pythia8+Delphes, single-charm-jet final-state]

At LHC, EIC, etc. jets have common features…



Calorimeter activity (charged + neutral particles)

At LHC, EIC, etc. jets/HF have common features…



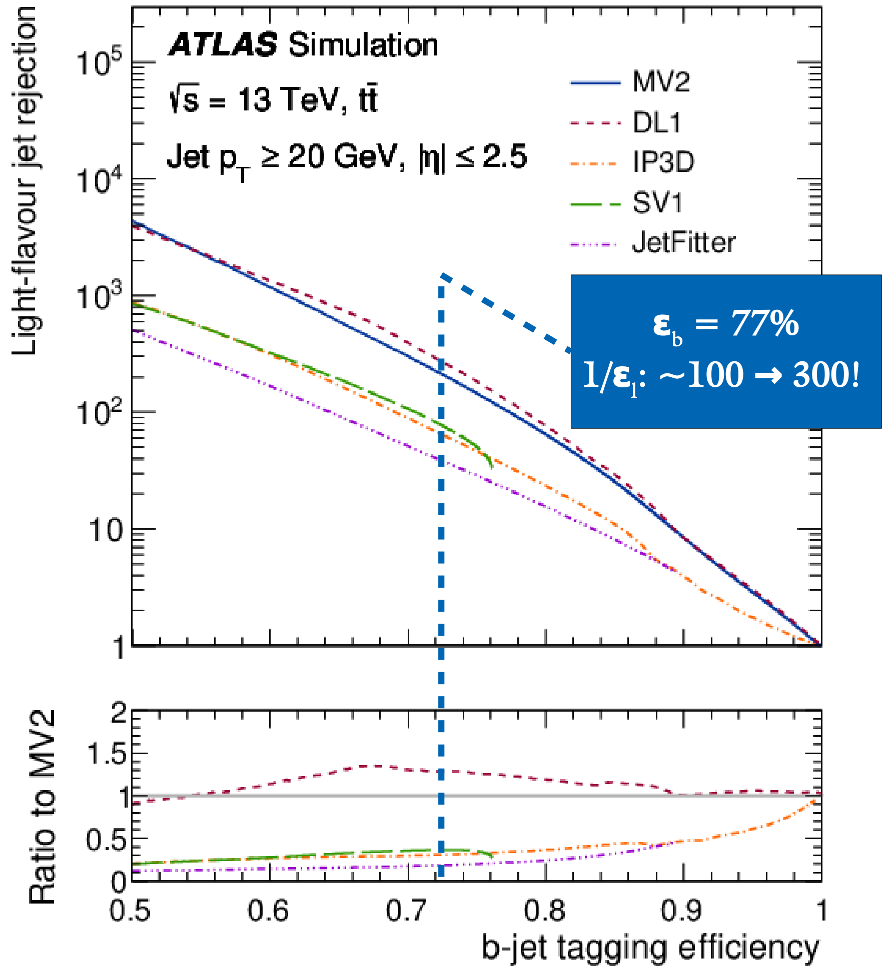
Hadron	cτ [mm]
$B^{0,\pm}$	~0.450
D^0	0.123
$D^\pm$	0.312
K_s^0	2.7
Λ	7.9

Targets for AI/ML

- Calorimetry [c.f. D. Romanov, "AI for Calorimetry"]
  - improving clustering, calibration, etc. → all benefit large-scale questions like "is this a heavy-flavor decay/heavy-flavor jet"?
- Tracking [c.f. L.-G. Gagdon, "ML for tracking in HEP" and G. Gavalian, "AI for tracking at JLAB"]
  - improving hit splitting, fake track rejection, etc. → crucial benefit to refining track selection for eventual jet/heavy-flavor identification
- My focus
  - **AI/ML for combining low-level or high-level track/calorimeter information with intended purpose of identifying jets and/or heavy flavor states**
  - Focus on examples/lessons from LHC experiments, while recognizing the challenges at LHC are not the same as those at EIC → nevertheless, the gains for this application seen at LHC should provide insights for major strides at EIC experiments!

ML/AI from LHC Run 1 to Run 2: Big Picture

Progression of Jet/HF identification with time and methods

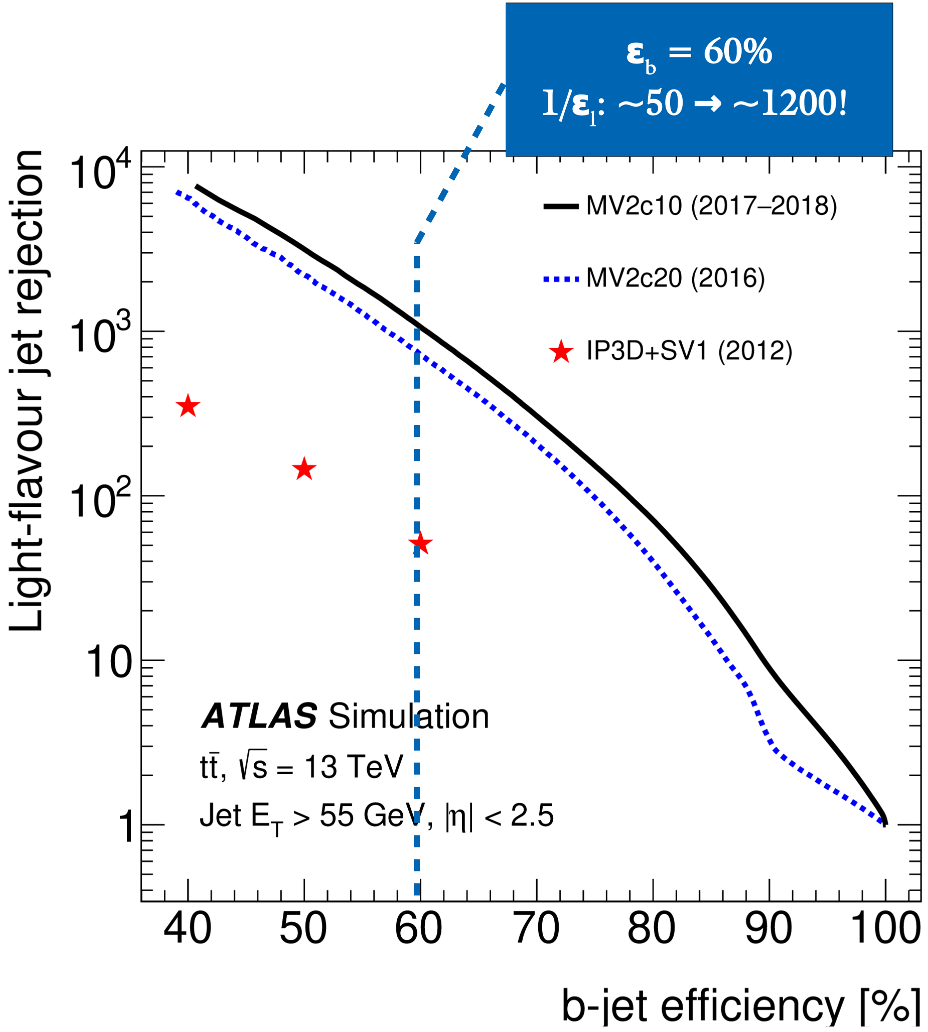


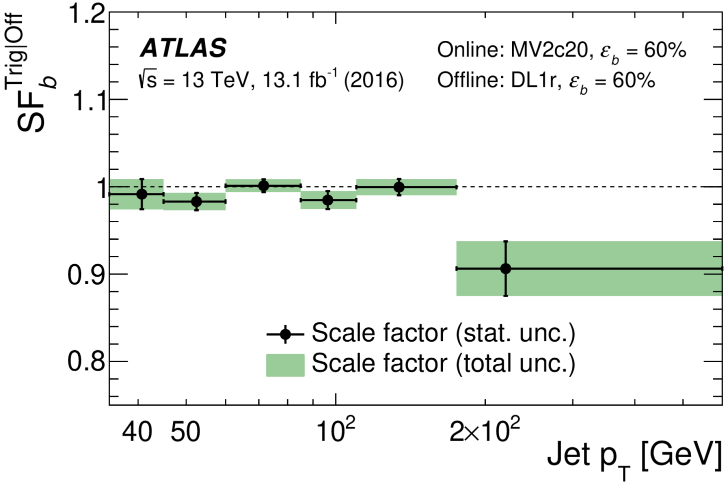
SV1, IP3D, JetFitter: dominant methods in LHC Run 1 and very early Run 2 (2010-2015) → single-feature-based approaches

MV1: dominant method in Run 2 (2016-2018) → multi-feature ML discriminant (boosted decision tree)

DL1: final Run 2 and early Run 3 (2020-present) method → multi-feature deep-learning discriminant.

… and on online/real-time applications, too!

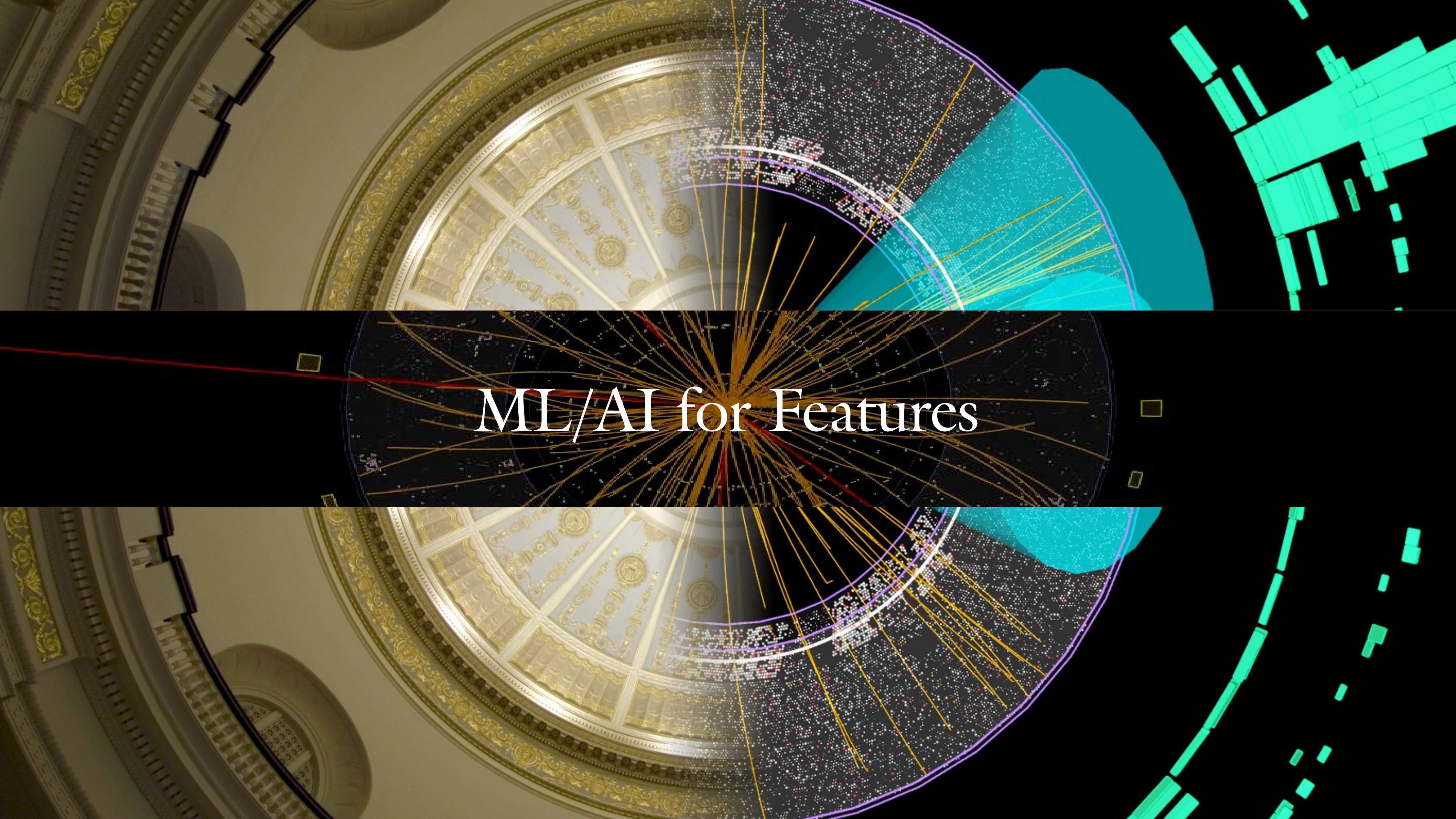




Online/trigger applications always lag offline applications due to more conservative nature of operations. Nevertheless, experiments moved as swiftly as possible to implement offline approaches in online applications, and to achieve high-fidelity performance compared to offline → *reduced trigger-related systematic uncertainties.*

At highlighted working point, online and offline performance agree to within about 5-10%.

ML/AI for Features



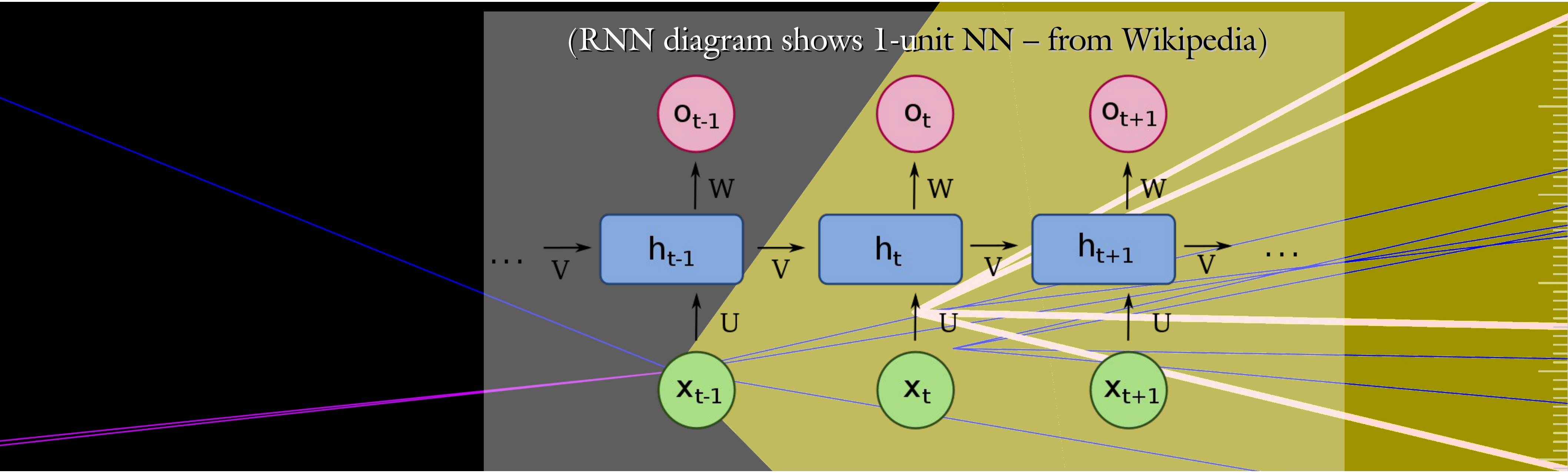
Recurrent Neural Network for Space-Time Sequences

A real heavy-flavor decay is a sequence of correlated events in space-time

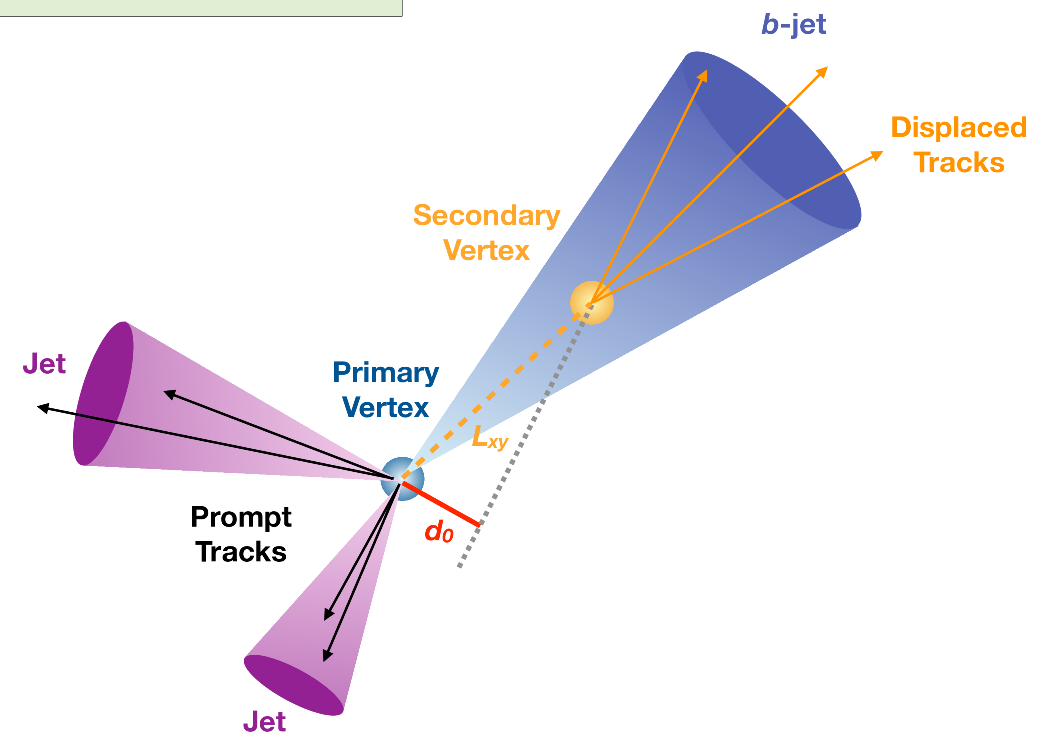
Light-flavor decays are generally more prompt and sequences are coincidences.

Recurrent Neural Networks (RNNs) are designed exactly to learn about sequence-based or time-ordered domains.

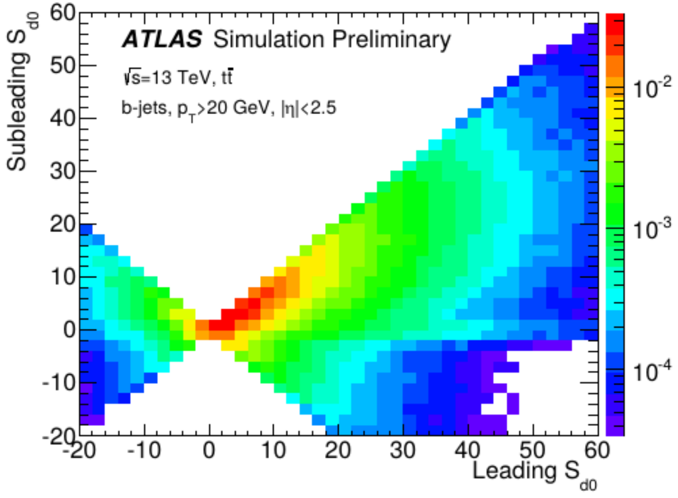
(RNN diagram shows 1-unit NN – from Wikipedia)



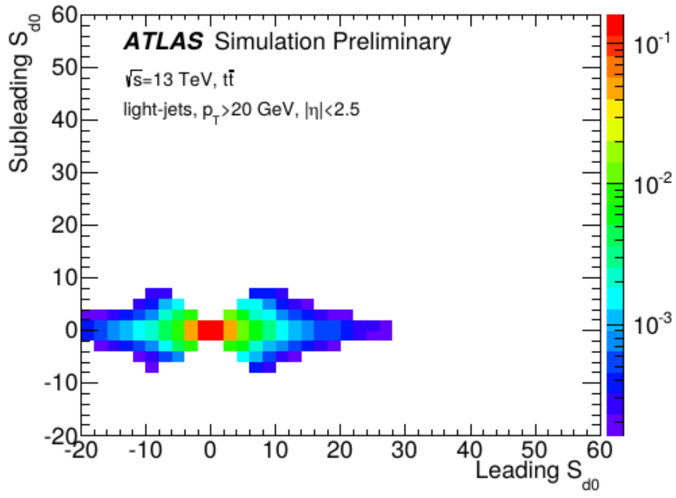
arXiv:2106.03584
ATL-PHYS-PUB-2017-003



b



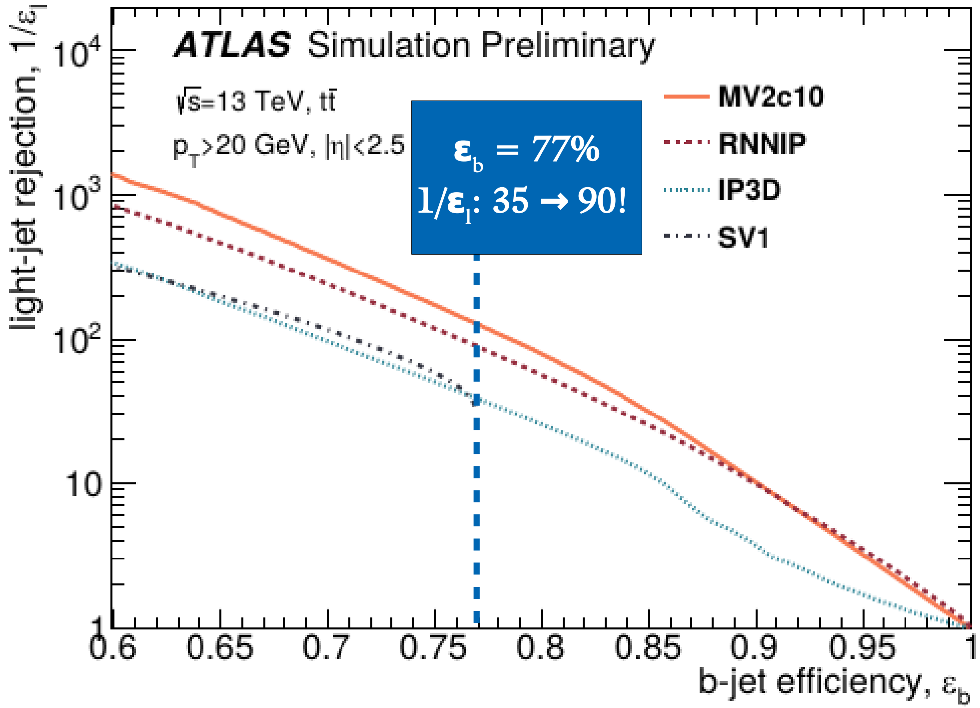
light

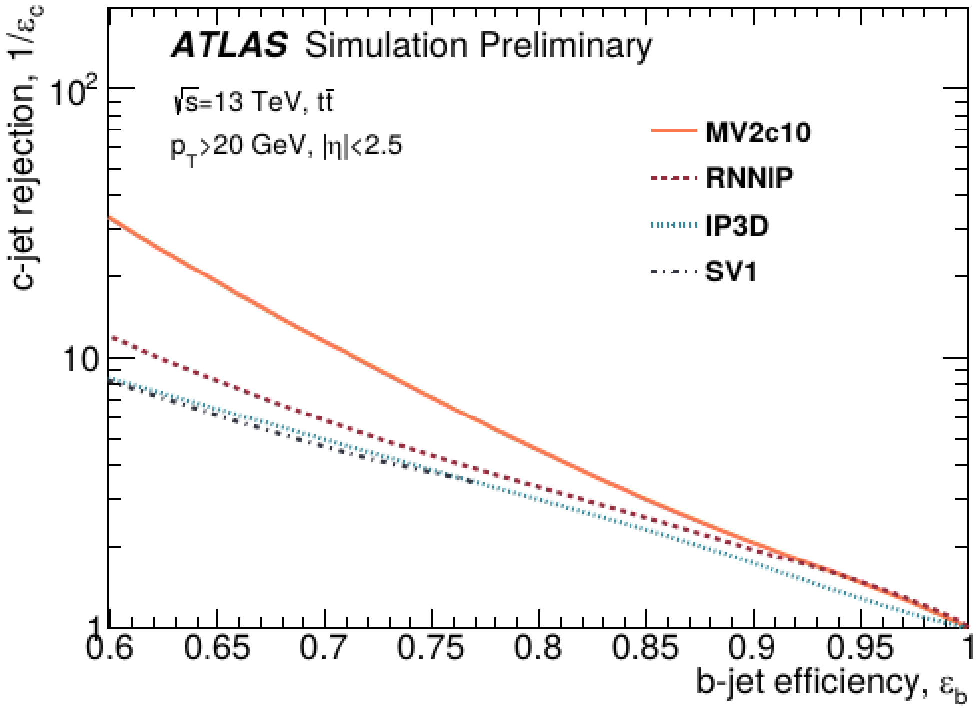


2-D and 3-D impact parameters (e.g. d_0) a useful measure ("Feature") for presence of "displaced tracks" (sign information comes from dot product of jet vector with track IP-POCA/DOCA vector.)

$$D_{\mathrm{RNN}} = \ln \frac{p_b}{f_c p_c + f_\tau p_\tau + (1 - f_c - f_\tau) p_{\mathrm{light}}}$$

Train RNN on track flight significance and momentum/angle relationships in jet, and define a likelihood score from the outputs (one for each category)





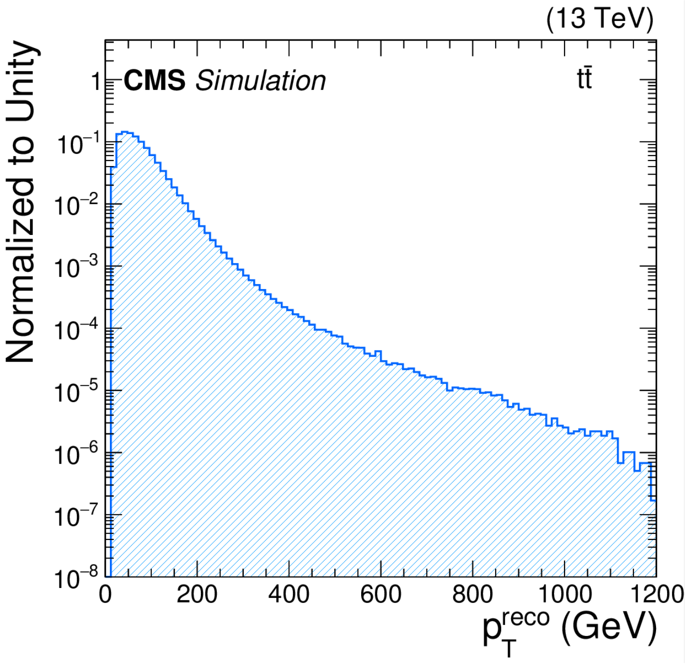
IP3D = original likelihood-based approach using only impact parameters; MV2c10 = multi-feature boosted decision tree discriminant; SV1 = secondary vertex-only tagging

RNNIP exceeds performance of other "single-feature" taggers, even when trained on same variables.

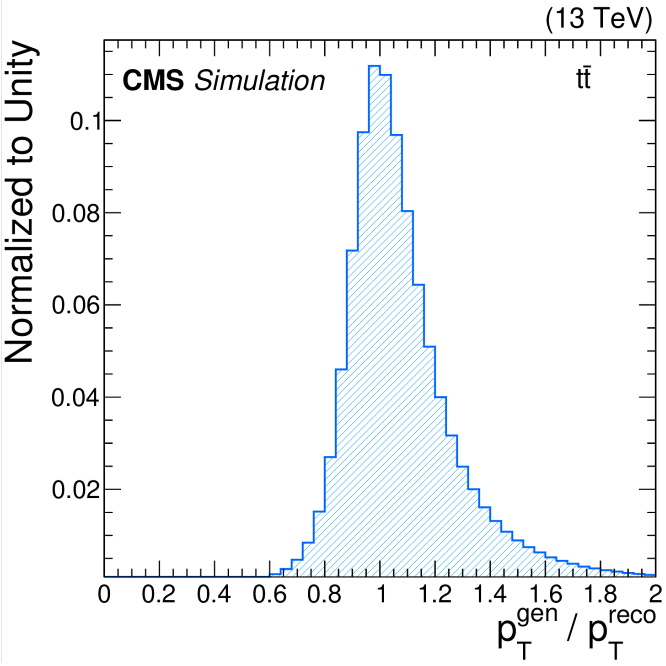
b-jet Energy and Resolution

arXiv:1912.06046

Heavy flavor jets produce more charged leptons and neutrinos → Dedicated corrections are needed especially for these jets.



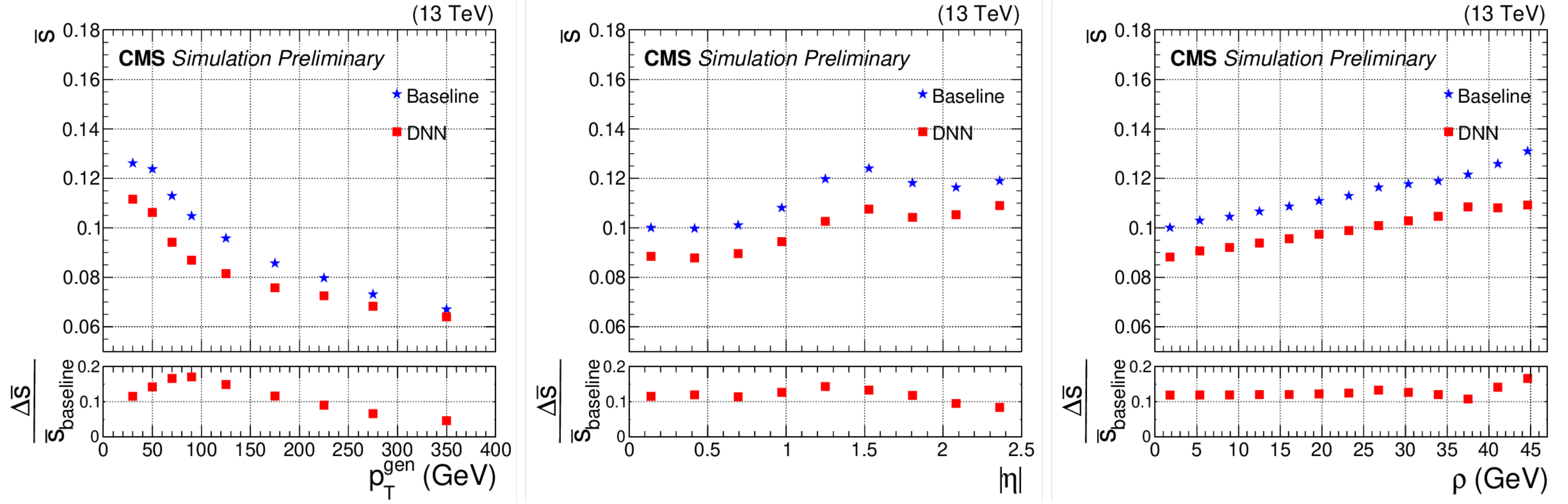
Reconstructed p_T



Regression target

Fidelity to "true value" has a strong benefit in turning experimental results back into fundamental statements.

Train a deep neural network on jet kinematic, event pileup, leptons matched to jets, vertexing, and jet constituent (e.g. leading constituents) information. Use Huber loss function with three output targets: the mean estimator and the 25% and 75% quantiles of the target regression distribution.



MC sample	Improvement
$t\bar{t}$	12.2%
$Z(\rightarrow \ell^+\ell^-)H(\rightarrow b\bar{b})$	12.8%
$H(\rightarrow b\bar{b})H(\rightarrow \gamma\gamma)$ SM	13.1%
$H(\rightarrow b\bar{b})H(\rightarrow \gamma\gamma)$ resonant 500 GeV	14.5%
$H(\rightarrow b\bar{b})H(\rightarrow \gamma\gamma)$ resonant 700 GeV	13.1%

Significant gains in jet-energy resolution with relatively similar performance across a range of b-jet analysis applications.

arXiv:1912.06046

ML/AI for Whole-Object Identification

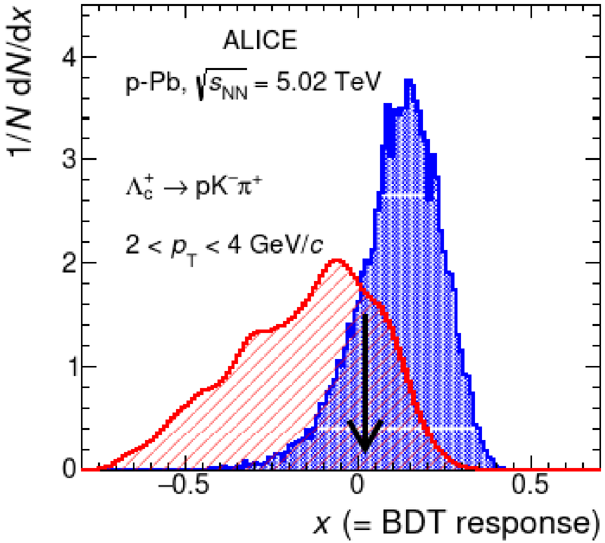
AI/ML for Heavy Flavor States

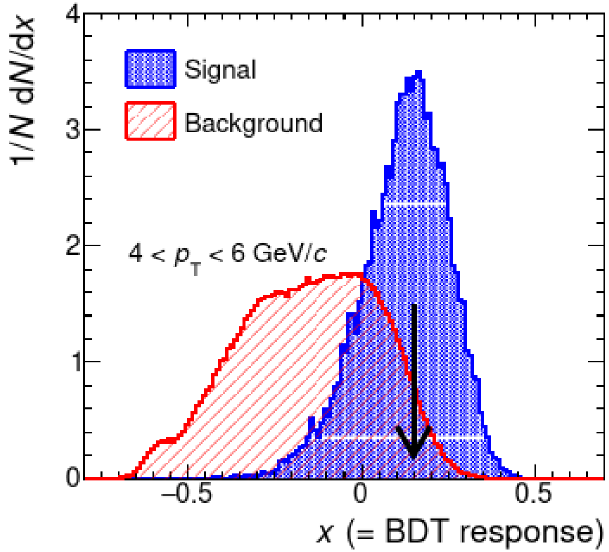
The use of machine learning for identifying heavy flavor states is well-documented and widespread → *"keep on keeping on"*

Λ^{+}_{c} production in p-Pb collisions

Important test of pQCD and for informing future calculations, understanding Quark-Gluon Plasma, etc.

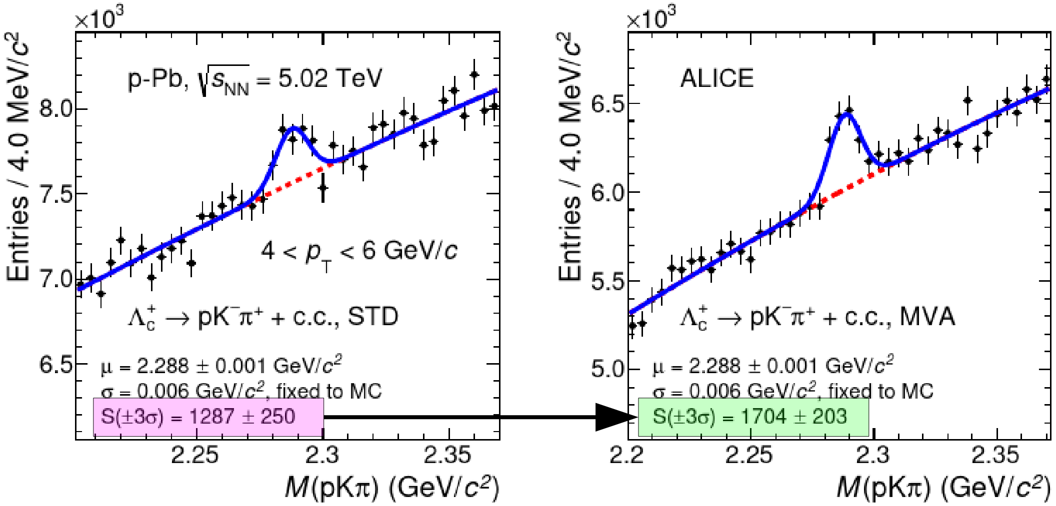
Train on kinematic information about daughter tracks, PID information, and decay length information.



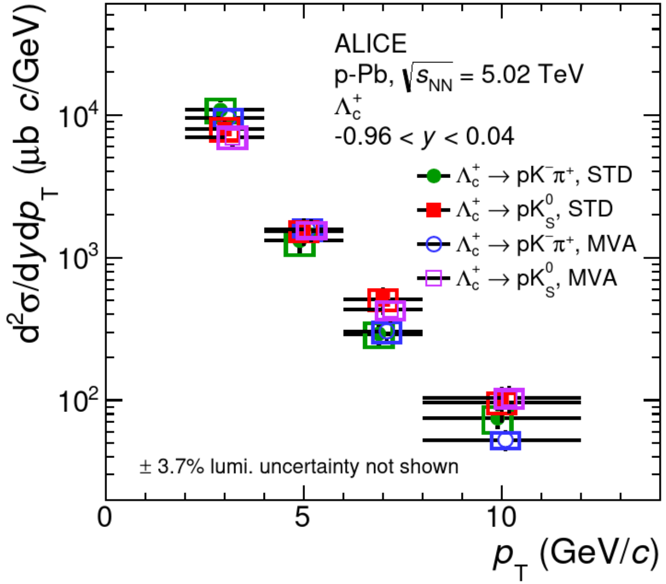


AI/ML for Heavy Flavor States

arXiv:2106.08278
arXiv:1712.09581



	$\Lambda_c^+ \to pK^-\pi^+$				
	STD		MVA		
	lowest p_T	highest p_T	lowest p_T	highest p_T	l
Yield extraction (%)	10	11	7	4	
Tracking efficiency (%)	10	7	10	7	
Cut efficiency (%)	9	12	8	6	
PID efficiency (%)	6	6	neg.	neg.	
MC p_T shape (%)	2	2	neg.	3	
Multiplicity (%)	neg.	neg.	neg.	neg.	
Beauty feed-down (%)	$^{+1}_{-5}$	$^{+2}_{-10}$	$^{+1}_{-5}$	$^{+2}_{-10}$	
Branching ratio (%)	5.1				
Luminosity (%)					3.5



The multivariate analysis (MVA) is consistently improved over the standard (STD) analysis, both in statistical and in the impact of systematic uncertainties. MVA approach leads to overall improved precision with same data.

Deep-Learning Many-Feature/High-Level Approaches

Performance improvements achieved are notable and significant (described earlier in the talk, e.g. 3x or greater improvement in light-jet rejection).

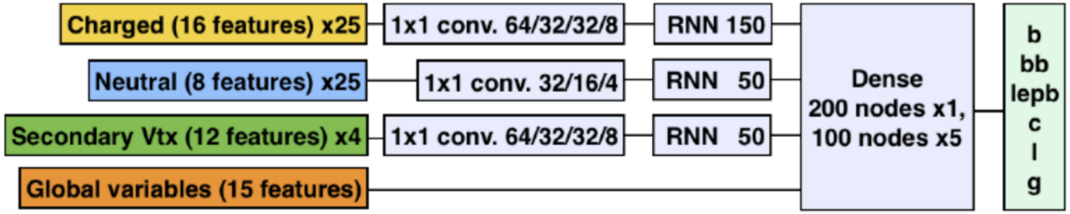
Focus here will be on validation of these methods using data and simulation, which from an experimental perspective is the most important factor for any method, ML/AI or not.

ATLAS "DL1"

$$D_{\mathrm{DL1}} = \ln\left(\frac{p_b}{f_c \cdot p_c + (1 - f_c) \cdot p_{\mathrm{light}}}\right)$$

Input	Variable	Description
Kinematics	p_{T}	Jet p_{T}
	η	Jet $\lvert\eta\rvert$
IP2D/IP3D	$\log(P_b/P_{\mathrm{light}})$	Likelihood ratio between the b-jet and light-flavour jet hypotheses
	$\log(P_b/P_{\mathrm{c}})$	Likelihood ratio between the b- and c-jet hypotheses
	$\log(P_c/P_{\mathrm{light}})$	Likelihood ratio between the c-jet and light-flavour jet hypotheses
SV1	$m(\mathrm{SV})$	Invariant mass of tracks at the secondary vertex assuming pion mass
	$f_E(\mathrm{SV})$	Energy fraction of the tracks associated with the secondary vertex
	$N_{\mathrm{TrkAtVtx}}(\mathrm{SV})$	Number of tracks used in the secondary vertex
	$N_{\mathrm{2TrkVtx}}(\mathrm{SV})$	Number of two-track vertex candidates
	$L_{xy}(\mathrm{SV})$	Transverse distance between the primary and secondary vertex
	$L_{xyz}(\mathrm{SV})$	Distance between the primary and the secondary vertex
	$S_{xyz}(\mathrm{SV})$	Distance between the primary and the secondary vertex divided by its uncertainty
	$\Delta R(\vec{p}_{\mathrm{jet}}, \vec{p}_{\mathrm{vtx}})(\mathrm{SV})$	ΔR between the jet axis and the direction of the secondary vertex relative to the primary vertex.
JetFitter	$m(\mathrm{JF})$	Invariant mass of tracks from displaced vertices
	$f_E(\mathrm{JF})$	Energy fraction of the tracks associated with the displaced vertices
	$\Delta R(\vec{p}_{\mathrm{jet}}, \vec{p}_{\mathrm{vtx}})(\mathrm{JF})$	ΔR between the jet axis and the vectorial sum of momenta of all tracks attached to displaced vertices
	$S_{xyz}(\mathrm{JF})$	Significance of the average distance between PV and displaced vertices
	$N_{\mathrm{TrkAtVtx}}(\mathrm{JF})$	Number of tracks from multi-prong displaced vertices
	$N_{\mathrm{2TrkVtx}}(\mathrm{JF})$	Number of two-track vertex candidates (prior to decay chain fit)
	$N_{\mathrm{1\text{-}trk\ vertices}}(\mathrm{JF})$	Number of single-prong displaced vertices
	$N_{\geq 2\text{-}\mathrm{trk\ vertices}}(\mathrm{JF})$	Number of multi-prong displaced vertices
JetFitter c-tagging	$L_{xyz}(2^{\mathrm{nd}}/3^{\mathrm{rd}}\mathrm{vtx})(\mathrm{JF})$	Distance of 2^{nd} or 3^{rd} vertex from PV
	$L_{xy}(2^{\mathrm{nd}}/3^{\mathrm{rd}}\mathrm{vtx})(\mathrm{JF})$	Transverse displacement of the 2^{nd} or 3^{rd} vertex
	$m_{\mathrm{Trk}}(2^{\mathrm{nd}}/3^{\mathrm{rd}}\mathrm{vtx})(\mathrm{JF})$	Invariant mass of tracks associated with 2^{nd} or 3^{rd} vertex
	$E_{\mathrm{Trk}}(2^{\mathrm{nd}}/3^{\mathrm{rd}}\mathrm{vtx})(\mathrm{JF})$	Energy fraction of the tracks associated with 2^{nd} or 3^{rd} vertex
	$f_E(2^{\mathrm{nd}}/3^{\mathrm{rd}}\mathrm{vtx})(\mathrm{JF})$	Fraction of charged jet energy in 2^{nd} or 3^{rd} vertex
	$N_{\mathrm{TrkAtVtx}}(2^{\mathrm{nd}}/3^{\mathrm{rd}}\mathrm{vtx})(\mathrm{JF})$	Number of tracks associated with 2^{nd} or 3^{rd} vertex
	$Y_{\mathrm{trk}}^{\mathrm{min}}, Y_{\mathrm{trk}}^{\mathrm{max}}, Y_{\mathrm{trk}}^{\mathrm{avg}}(2^{\mathrm{nd}}/3^{\mathrm{rd}}\mathrm{vtx})(\mathrm{JF})$	Min., max. and avg. track rapidity of tracks at 2^{nd} or 3^{rd} vertex

CMS "DEEPJET"



Performance Assessment in Data: Examples

APPROACH

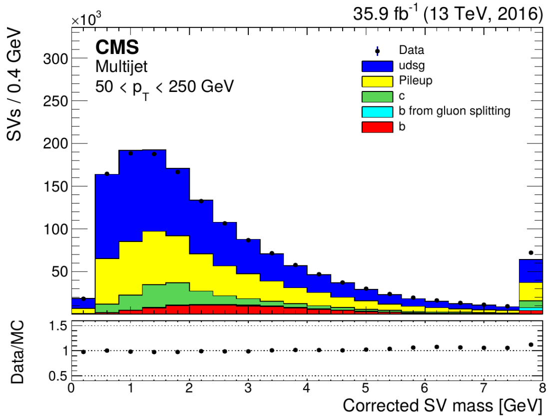
Multijet Events
(Trigger Prescaled)

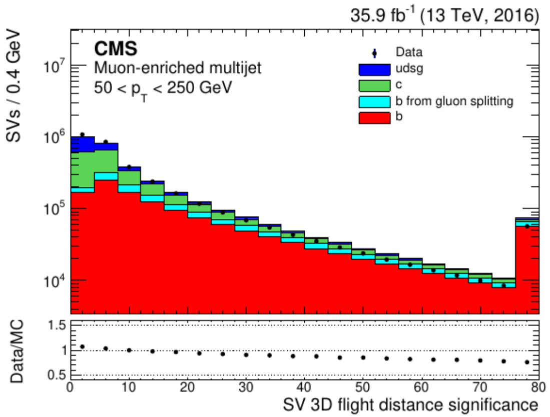
BACKGROUND-DOMINATED

Muon-Enriched Multijet
(Trigger Prescaled)

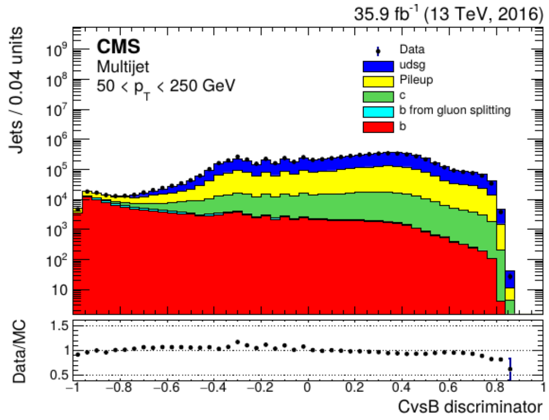
SIGNAL-ENRICHED

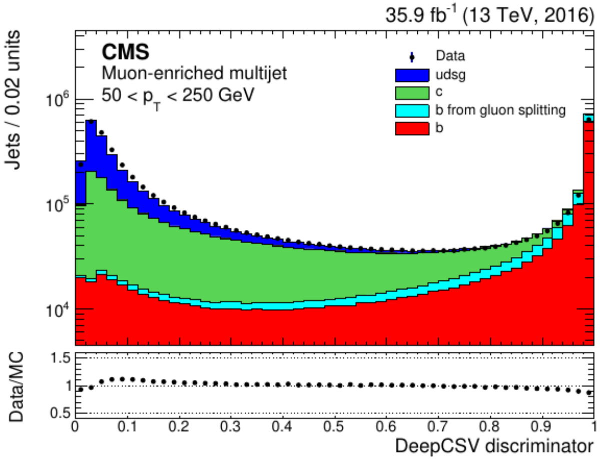
FEATURE EXAMPLE





TAGGER EXAMPLE





2018 JINST 13 P05011

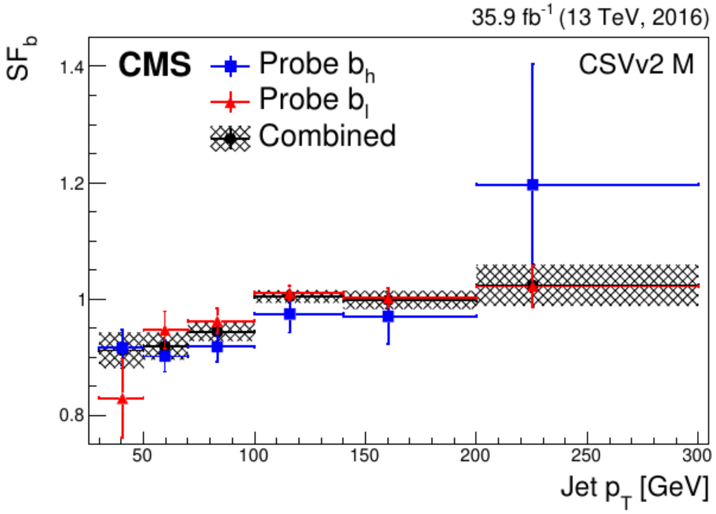
Data/MC Simulation Corrections

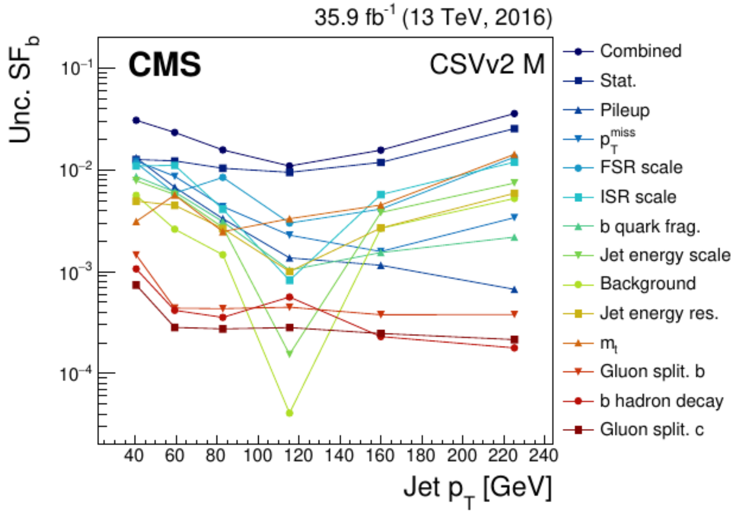
2018 JINST 13 P05011

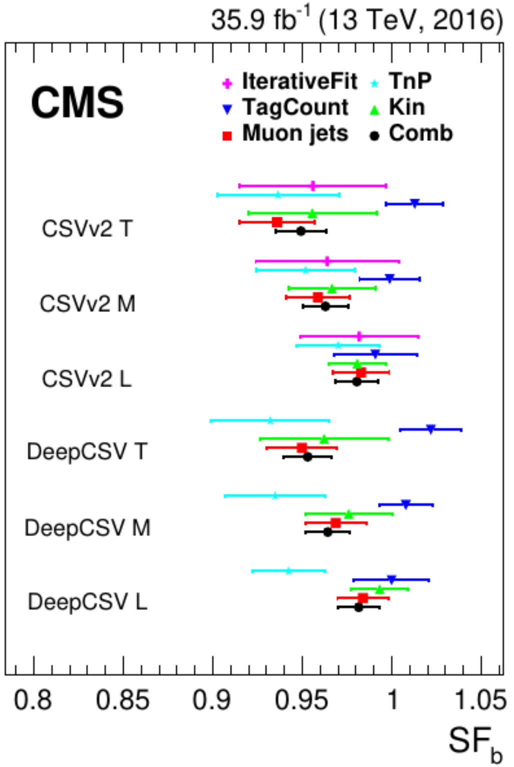
Using control-region data and simulation, define a per jet data-to-simulation “scale factor” (SF):

$$SF_f = \varepsilon_f^{\mathrm{data}}(p_{\mathrm{T}}, \eta) / \varepsilon_f^{\mathrm{MC}}(p_{\mathrm{T}}, \eta)$$

Example using single-lepton top-pair event selection







Simulation approaches are within 5-10% of the data → expect that to get better.

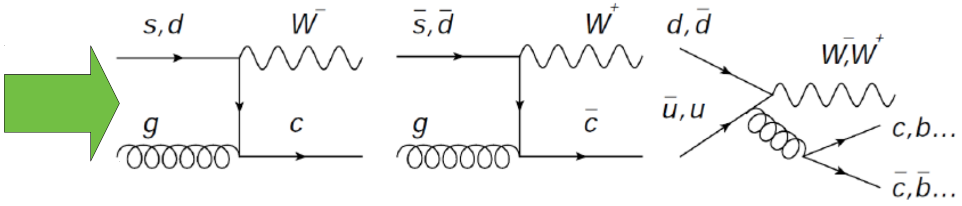
Charm Tagging

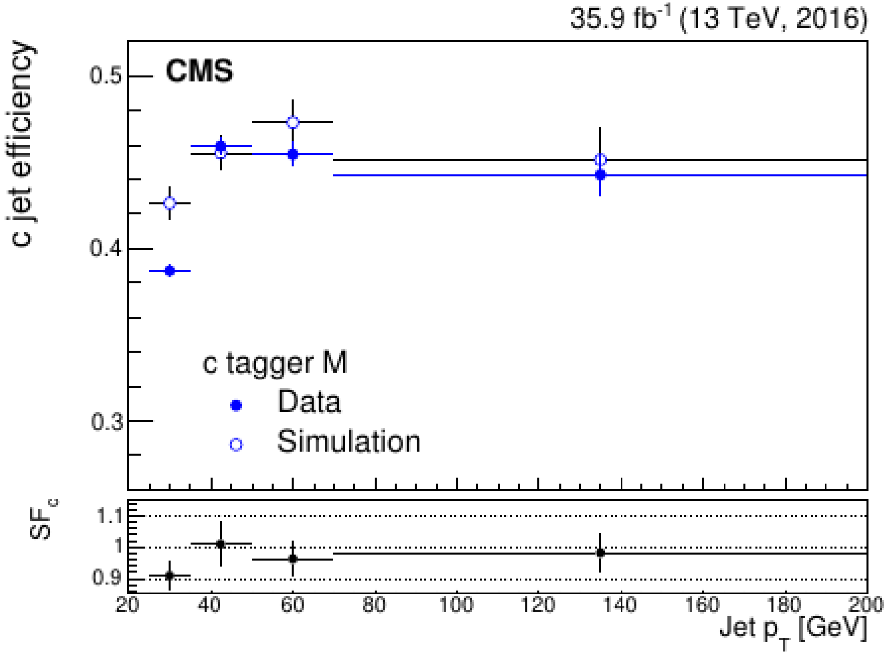
2018 JINST 13 P05011
arXiv:1802.04329
ATL-PHYS-PUB-2020-009

The ML/AI era has enabled advanced approaches not just to b-jets but to more-difficult-to-tag charm jets, which are definitely “heavy flavor” but more similar to light jets than are b-jets.

EXAMPLE: W+c events

(I like these because of the extreme similarity to CC DIS at EIC…!)

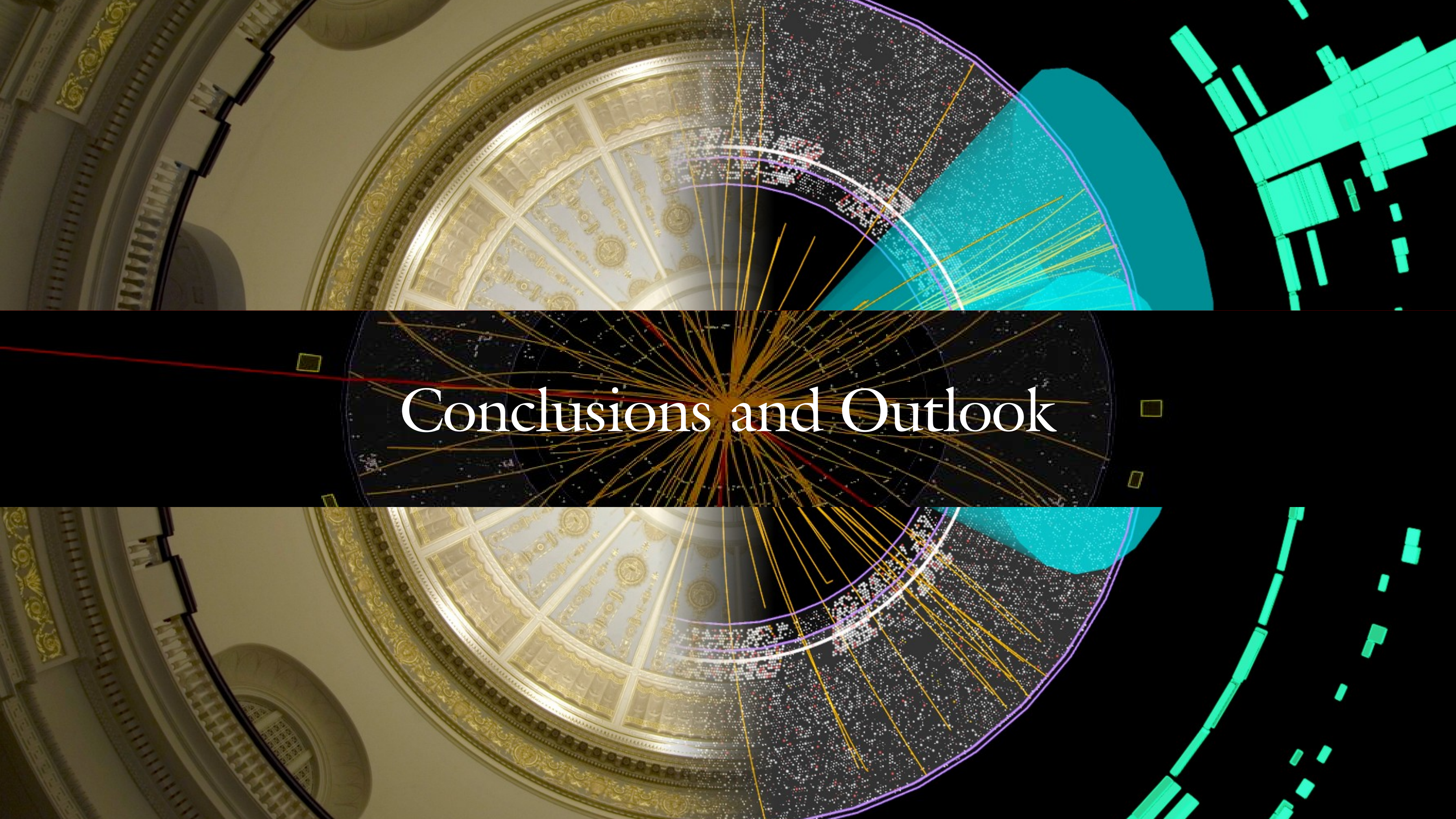




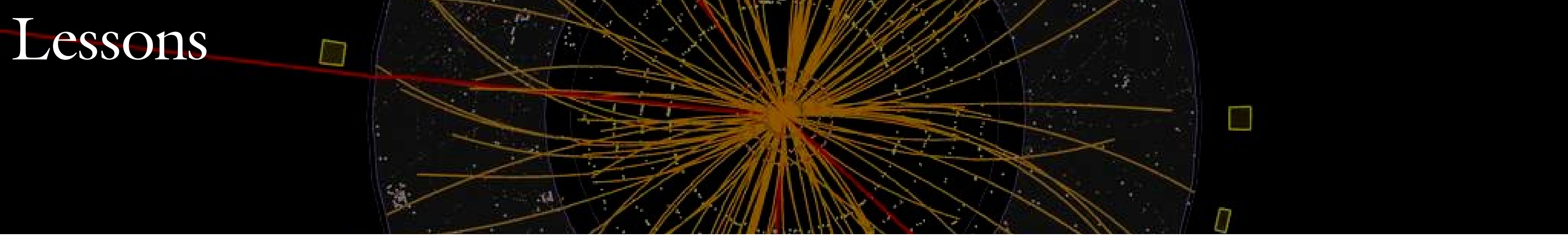
CMS (and ATLAS) have similar conceptual approaches (DNN, BDTs, etc.) and each reports a light mis-identification rate ranging from about 1%-5% across a momentum range spanning 1000 GeV, for a charm jet efficiency of ~40%.

Adopting similar approaches, long-term, for the EIC should bring great benefits here!

Conclusions and Outlook

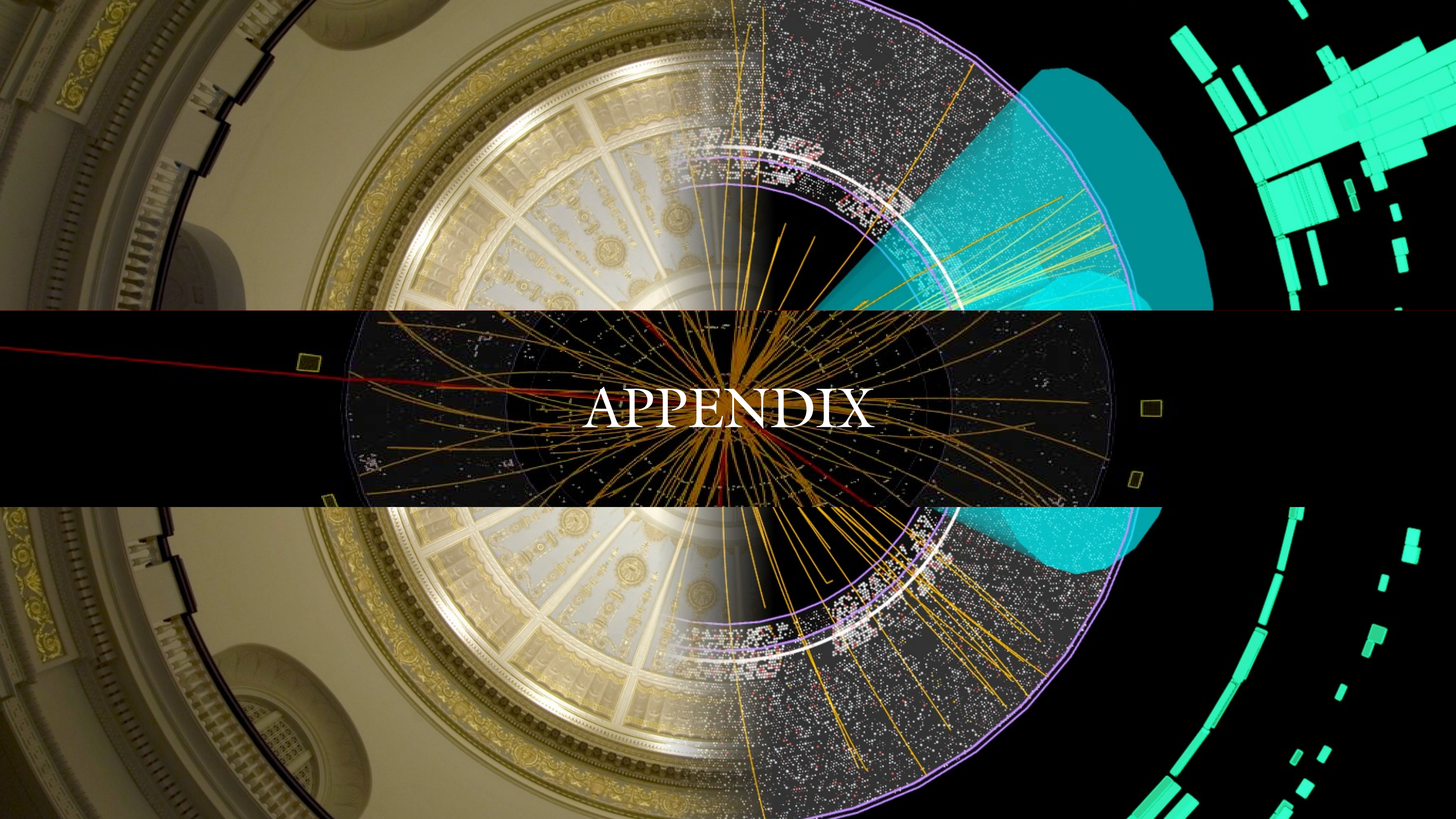


Lessons



- Deep-Learning approaches can be superior, even given the same (limited) information as earlier approaches (BDT, likelihood ratios, etc.)
  - Be attentive to fidelity between online/real-time application of approach and offline application → reduce systematics
- Modern simulation techniques can provide reliable training samples, but caution is nevertheless always warranted
  - For example: data/MC correction factors ("scale factors") not enlarged by using more information with deeper learning methodologies, despite potential risks of using lots of deep information whose modeling may not be as reliable as the whole.
- Validation, validation, validation: trust is built, as always, by assessing performance in as many ways (ideally on real data) as possible. Trust in the application of these advanced methods to places where we cannot yet check performance is built of trust in performance where we already knew/know the answer.
  - For example: higher acceptance, better resolution, etc. allow for phase space/physical space in detector experiments to be carved up into many control regions → more validation with more controls.

APPENDIX

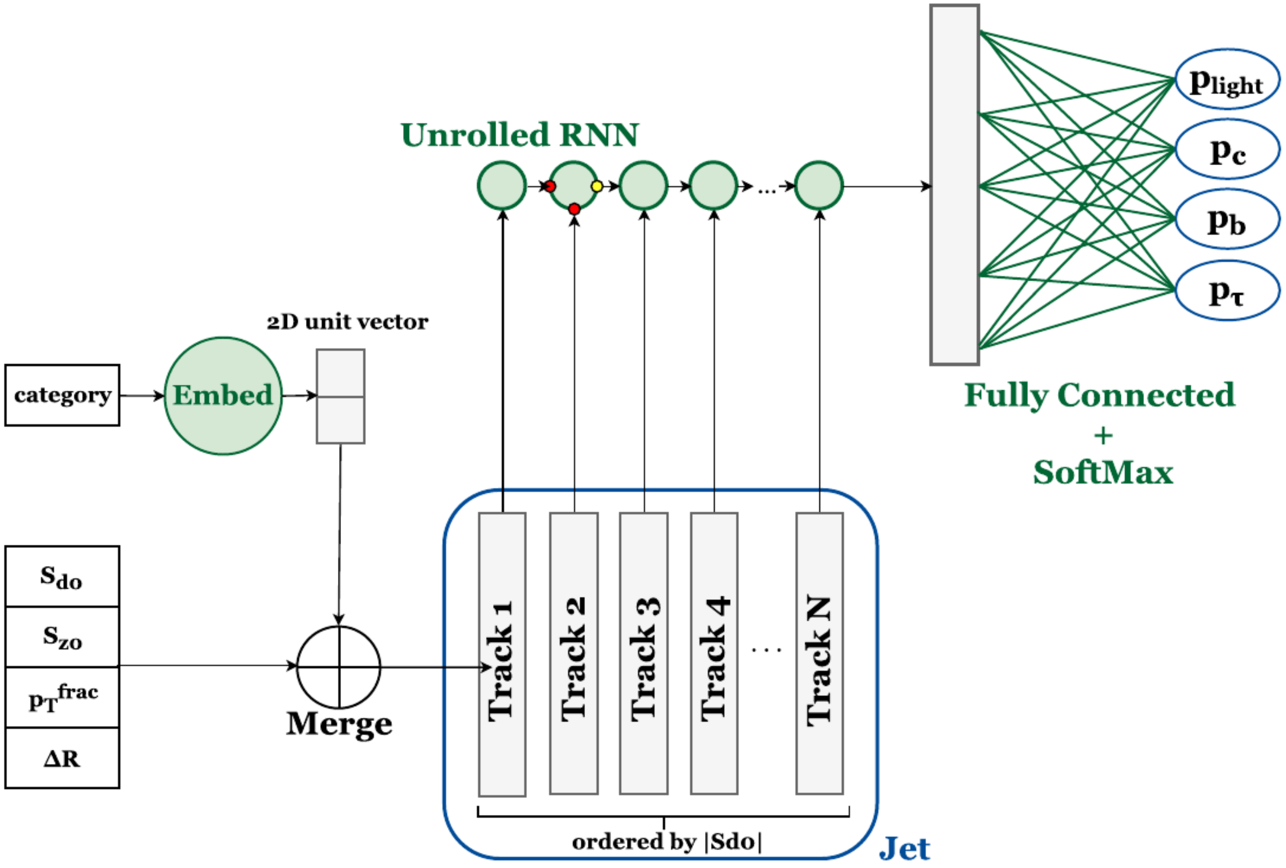


LHC Running Schedule

https://lhc-commissioning.web.cern.ch/schedule/LHC-long-term.htm

ATLAS RNN Tagger Architecture

ATL-PHYS-PUB-2017-003



ATLAS “DL1(r)” Architecture

ATL-PHYS-PUB-2017-013

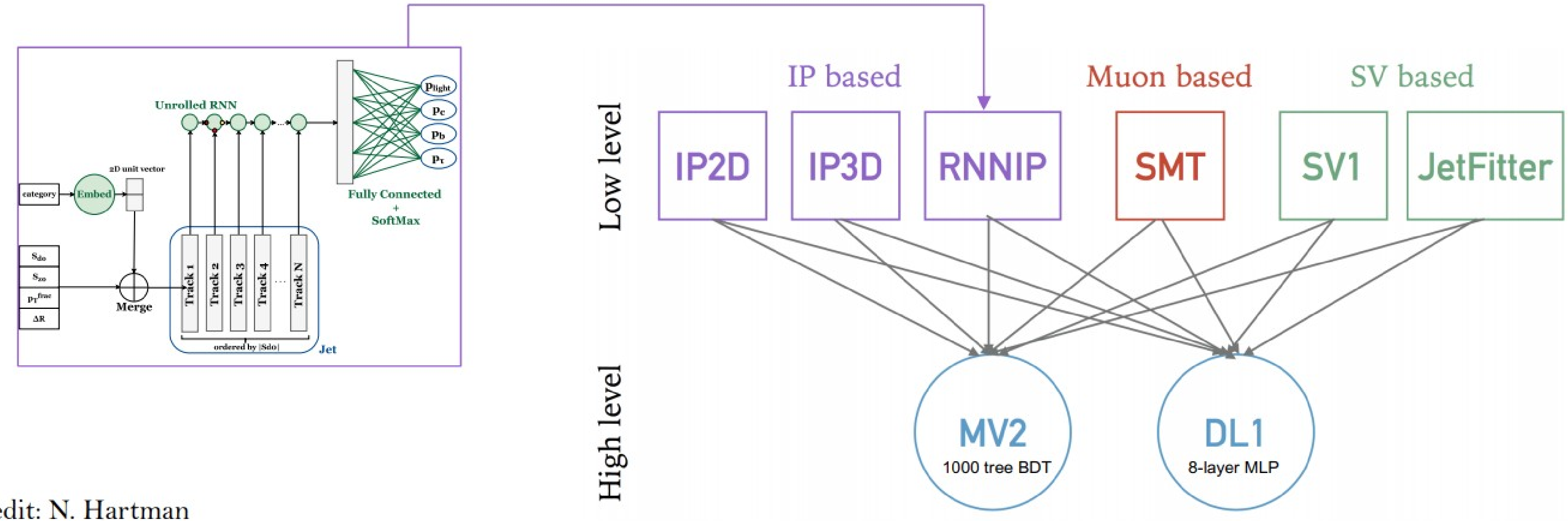
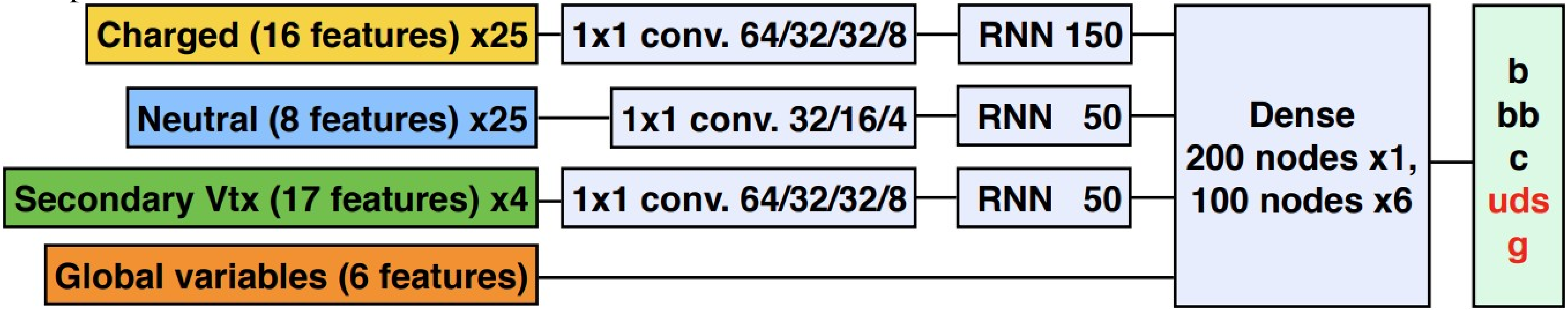


Image credit: N. Hartman

CMS “DeepJet”

CMS-DP-2017-013
2018JINST13P05011

Employs KERAS + TensorFlow (low-level operations like convolution) for training for “Combined Secondary Vertex Tagger” (CSV) → this and other feature taggers are combined into a single BDT output called “cMVAv2”.



DeepJet: A deep-neural-network algorithm based on 18 properties of up to 25 charged and 6 properties of 25 neutral particle-flow jet constituents, as well as 12 properties from up to 4 secondary vertices associated with the jet. For each collection of charged and neutral particles and vertices, separate 1x1 convolutional layers are trained: 4 hidden layers with 64,32,32, and 8 filters for charged candidates and vertices and 3 hidden layers with 32,16, and 4 filters for neutral particles. The filters act on each particle or vertex individually. The compressed and transformed output is fed through a separate recurrent layer for each collection with 150 nodes for charged candidates and 50 nodes for neutral candidates and secondary vertices. The output of these layers combined with global variables such as pT and η of the jet and is further processed by one dense layer with 200 nodes, followed by 7 hidden dense layers each with 100 nodes